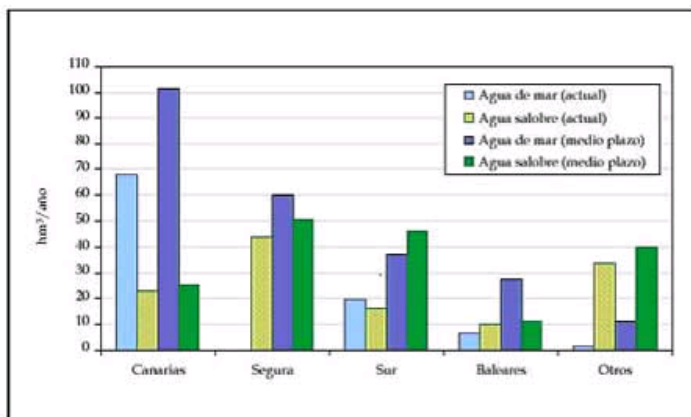


## LA ESTADÍSTICA NOS AYUDA

Cuando vemos en televisión **noticias relacionadas con el medio ambiente**, o cuando las leemos en los periódicos o en las revistas, casi siempre vienen acompañadas de **gráficos** extraños...



de cifras...

Año	Total anual remonte salmones	Total anual remonte multi-invierno	Total anual pesca salmones	Total anual pesca multi-invierno	Total antes del 15 de Junio pesca multi-invierno (y % del total)
1989	Sin datos	Sin datos	16	4	2 (50%)
1990	Sin datos	Sin datos	27	11	7 (64%)
1991	Sin datos	Sin datos	22	19	13 (68%)
1992	903	58	59	5	4 (80%)
1993	379	45	59	13	10 (77%)
1994	373	85	59	27	25 (93%)
1995	478	42	49	16	11 (69%)
1996	240	96	47	32	28 (87%)
1997	224	35	39	18	10 (56%)
1998	263	20	36	10	5 (50%)
1999	128	21	21	11	10 (91%)
2000	219	54	35	21	15 (71%)
2001	334	61	69	34	31 (91%)
2002	467	47	44	14	10 (71%)
2003	197	150	44	40	30 (75%)
2004	328	60	48	22	15 (68%)
2005	276	49	37	26	22 (85%)

y de **afirmaciones** tan rotundas como éstas que puedes ver aquí:

- (PD/Agencia EFE).- *El calentamiento global ha resquebrajado las plataformas de hielo en la Antártida y el aumento de icebergs ha alterado los sistemas ecológicos en torno a esas moles de hielo, según un estudio publicado en la revista Science.*
- (PD/Agencias).- *El Fondo Mundial para la Naturaleza (WWF, por sus siglas en inglés) critica en un informe la "frenética construcción" de desalinizadoras en España y su impacto negativo en el medioambiente y el cambio climático.*



Son muy importantes los **estudios estadísticos** que **conducen a este tipo de informaciones**, ya que basándose en estos estudios los gobiernos diseñan sus planes de gestión del medio ambiente.

De hecho existe una asignatura ofertada en distintas Universidades españolas denominada "Modelos estadísticos para el medio ambiente" y empresas dedicadas a hacer estudios estadísticos sobre temas medioambientales.



Pero **un estudio estadístico no se hace de la noche a la mañana**; es algo que requiere una **cuidadosa planificación**. Por ejemplo si se quiere conocer la población de aves en Doñana, el nivel de contaminación de los acuíferos o cómo controlar los residuos urbanos, necesitaremos:

- conocer determinados datos objetivos
- ordenarlos
- analizarlos
- sacar conclusiones.

Todo esto lo vas a hacer tú en este tema. No te preocupes, no será muy difícil, aunque tendrás que hacer cuentas...

Esta última parte, la de sacar conclusiones, es una de las más importantes. Se necesita **desarrollar un espíritu crítico** ante las noticias basadas en estudios



estadísticos, que se dan como "verdades absolutas". Es más frecuente de lo que parece el que datos objetivos sean interpretados según interese a quien los muestra.

Comprueba con la siguiente autoevaluación si tienes un buen "ojo crítico" a la hora de analizar informaciones basadas en "estudios estadísticos".

### Comprueba que lo has entendido

1. ¿Cuál o cuáles de las siguientes informaciones te parecen claramente manipuladas o erróneas?

- Según un estudio estadístico, realizado a dos personas en un club náutico, se determina que a todos los españoles les encanta el buceo deportivo.
- Según un estudio estadístico, realizado por una compañía eléctrica, se sabe que los andaluces no aprecian que haya contaminación generada por las centrales térmicas en nuestro territorio.
- Un estudio estadístico determina que el cien por cien de los encuestados respiran cada día.



## Planificamos el trabajo: tomando los datos



Mientras Javier estaba con Luis en el parque leyó en el periódico una noticia que decía "Según un estudio estadístico, el 60,6% de los andaluces opina que principal problema relacionado con el medio ambiente son los incendios forestales".

Este titular le llevo a la siguiente reflexión "¿Cómo se puede saber lo que opinan los andaluces en general si yo soy de Jaén y a mí no me ha preguntado nadie?".

No le falta razón a Javier ¿verdad? Lo que él no sabe es que la estadística puede ayudarnos a obtener conclusiones que se pueden generalizar para un grupo muy grande de personas, sin necesidad de tener que preguntarle a todo el grupo. Pero para ello es necesario que el **estudio estadístico esté bien diseñado**.

Para realizar un estudio estadístico hay que empezar teniendo muy claro un par de cosas:

1. **¿Qué queremos saber?**
2. **¿De quién queremos saberlo?**

### ¿Qué queremos saber?

Muchos estudios estadísticos comienzan con una pregunta o preguntas sobre un tema concreto. En estos casos en primer lugar habrá que crear un **cuestionario**.



### Por ejemplo...

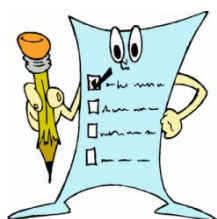
Si nos planteamos un estudio sobre "Impacto medioambiental en Andalucía" podríamos formular la pregunta:

"¿Qué problema relacionado con el medio ambiente le preocupa más?"



Las respuestas a esta pregunta pueden ser:

- **Abiertas:** cada persona puede dar tantas respuestas como le apetezca.
- **Abiertas pero limitadas:** cada persona entrevistada podría dar una o dos o tres o un número predeterminado de respuestas libres.
- **Cerradas:** cada persona entrevistada elige una o varias opciones sobre un listado prefijado de respuestas posibles



Por tanto, **habrá que decidir** si se crea un cuestionario:

- **Abierto:** cada uno puede contestar lo que quiera.
- **Limitado:** con un número prefijado de posibles respuestas.
- **Cerrado:** con un número determinado de respuestas que proporciona el propio cuestionario.

Este tipo de cuestionario (cerrado) es más cómodo para el entrevistado, pero puede "deformar" el estudio, ya que el listado de posibles respuestas va a depender del encuestador y su buen criterio. Para evitar la "manipulación" en un cuestionario cerrado, siempre debería existir la opción de respuesta "*otra respuesta diferente a las propuestas*".

Por ejemplo...



Observa en la siguiente animación algunos ejemplos de cuestionarios. Seguro que te aclararán las dudas sobre lo que es un cuestionario cerrado y un cuestionario limitado.

## ¿Cómo se realiza un cuestionario?



Queremos saber cuales serían las propuestas que haría un andaluz o andaluza para mejorar la situación medioambiental en Andalucía

Primer tipo de cuestionario: **CERRADO**

Al entrevistado se le ofrecen una serie de respuestas de entre las que debe elegir una o varias.

**Señale, entre estas opciones: ¿qué cree que se debería hacer para mejorar la situación medioambiental en Andalucía?**

- a. Nada, desde mi punto de vista la situación es inmejorable
- b. No me interesan los problemas medioambientales
- c. Estos problemas los deben resolver los políticos, que para eso les pagan
- d. Que lo piensen los pesados de los ecologistas, que no tienen otra cosa que hacer

Observa que **el cuestionario está bastante "dirigido"**, es decir, las opciones que se ofrecen están diseñadas de forma que la información que se obtenga tenga un signo conformista o negligente.

Esto ocurre frecuentemente en encuestas realizadas por medios de comunicación en las que las opciones que te proponen determinan el resultado.



### Segundo tipo de cuestionario: **ABIERTO PERO LIMITADO**

El entrevistado puede dar un número determinado de respuestas libres

Escriba tres respuestas posibles a la pregunta: ¿qué cree que se debería hacer para mejorar la situación medioambiental en Andalucía?

#### Entrevista: 1 *(soluciones "caseras")*

- a. Usar más el transporte público
- b. Controlar el consumo de electricidad en casa y el trabajo
- c. Reciclar los residuos (basura)

#### Entrevista: 2 *(soluciones "generales")*

- a. Hacer que se cumpla el protocolo de Kioto
- b. Presionar a las empresas y fábricas para que contaminen menos
- c. Crear productos no contaminantes

Este tipo de cuestionario **puede provocar el rechazo** del entrevistado/a, al ser más difícil de rellenar. También **complica el estudio estadístico posterior** ya que, probablemente, se deba trabajar con un número muy elevado de respuestas posibles.

### Tercer tipo de cuestionario: **CERRADO (con la opción de "otra respuesta")**

El entrevistado elige una respuesta de entre las propuestas o propone otra

Señale, entre estas opciones: ¿qué cree que se debería hacer para mejorar la situación medioambiental en Andalucía?

- a. Nada, desde mi punto de vista la situación es inmejorable
- b. No me interesan los problemas medioambientales
- c. Estos problemas los deben resolver los políticos, que para eso les pagan

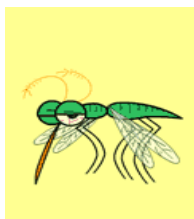
Para evitar la evidente "manipulación" que suponen estas respuestas, debemos añadir...

*Muchas cosas en nuestro día a día, como reciclar los residuos, usar más el transporte público, controlar el*

- a. Otra opción diferente *gasto de electricidad en el hogar, ...*

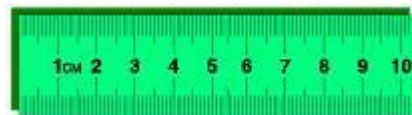
Quizá sea éste el mejor método para elaborar un cuestionario, ya que combina la comodidad para el entrevistado con las ventajas del "cuestionario abierto".

En el apartado de recursos del tema puedes ver el ejemplo anterior en una presentación llamada "*¿Cómo se elabora un cuestionario?*".



**No en todos los estudios estadísticos es necesario elaborar un cuestionario.** Si, por ejemplo, la pregunta que pretendemos responder fuese "¿Cuál es el tamaño en milímetros, de determinada especie de mosquito?", no necesitamos cuestionario.

Lo que necesitamos será un **instrumento de medida** adecuado que nos proporcione los **datos** que necesitamos. En las Ciencias, este



tipo de estudios estadísticos, en los que no es necesario preguntar, sino **medir**, es muy frecuente.



Al conjunto final de respuestas obtenidas le llamamos **VARIABLE ESTADÍSTICA**

Una variable estadística puede ser

CUALITATIVA	CUALITATIVA
Si no es un número	Si es un número
<p>Por ejemplo...</p> <ul style="list-style-type: none"> <li>Color preferido de un grupo de gente.</li> <li>Partido al que votarás en las próximas elecciones.</li> <li>Problema relacionado con el medio ambiente que preocupa más a los andaluces.</li> </ul>	<p>Por ejemplo...</p> <ul style="list-style-type: none"> <li>Altura de un grupo de personas.</li> <li>Gasto mensual de las familias de una ciudad en hipoteca.</li> <li>Peso de los pollos que se venden en los mercados andaluces</li> </ul>

En el caso de que el estudio se refiera, por ejemplo, a conocer "**el tamaño en milímetros de una determinada especie de mosquito**" no será necesario un cuestionario, habrá en este caso que elegir un sistema o **aparato de medida adecuado** para llevar a cabo las mediciones necesarias. Los datos obtenidos en este ejemplo concreto formarán una **variable cuantitativa**.

Comprueba que lo has entendido



2. Se quiere conocer la cantidad de CO<sub>2</sub> que hay en el aire en una determinada población. ¿Cuál sería la opción más adecuada para llevar a cabo este estudio?
  - a. Crear un cuestionario abierto preguntando por la cantidad de CO<sub>2</sub> que hay en el aire.
  - b. Crear un cuestionario cerrado con las respuestas: 20 mg/m<sup>3</sup>, 10 mg/m<sup>3</sup> y otra cantidad.
  - c. Instalar un aparato medidor en algún punto de la ciudad que registre los datos de cantidad de CO<sub>2</sub> que hay en el aire a lo largo de un periodo determinado de tiempo.
3. Indica si las siguientes variables aleatorias son cualitativas o cuantitativas:
  1. Energía aportada por distintas marcas de muesli.
  2. Sistema de calefacción utilizado en el invierno por familias de Madrid.
  3. Volumen de basura generado por las familias de una barriada de Granada.
  - a. Soluciones al problema de la contaminación de las aguas.



## ¿De quién lo queremos saber?

Imagínate que queremos saber cuál es el problema medioambiental que más preocupa a los vecinos de Villacañas de Abajo. Podría ocurrir que fuese muy distinto del que pueda preocupar a los vecinos de Villacañas de Arriba o a la población andaluza en su conjunto.

Por esto **es muy importante**, una vez terminado el cuestionario o método de recogida de datos adecuado al estudio, **decidir a quién va dirigido**:

- Si vamos a preguntar a todos los nacidos o nacidas en Andalucía o en alguna población en concreto de nuestra geografía
- Si se preguntaría a todas las personas censadas en Andalucía (o lugar en concreto de ésta)
- Si preguntaremos a cualquier persona que esté en ese momento dado en Andalucía (o en una determinada población andaluza)
- Etc.



Al **conjunto total de personas o de objetos de los que nos interesa conocer una determinada opinión o característica** es a lo que llamamos **población**.

Sea cual sea la elección, preguntar a toda la población normalmente es imposible, así que habrá que **elegir un grupo que represente toda la población**.



El **grupo elegido para que responda al cuestionario o del que se van a recoger determinados datos**, es a lo que se denomina **muestra**.



Cuanto mayor sea el número de personas de la muestra, más fiable será el estudio estadístico.

Y aquí es donde pueden empezar nuestros problemas porque  
**si elegimos mal la muestra los resultados no serán reales,  
 no serán fiables.**

¿Cómo elegir entonces la muestra?

Esto es una de las partes más complejas de la estadística y hay teorías matemáticas muy complicadas al respecto.

La elección de la muestra puede ser	Los dos métodos plantean problemas
<b>Aleatoria:</b> Se eligen al azar	La muestra puede no ser representativa del total de la población.
<b>Intencional:</b> El encuestador elige a los que quiere.	La muestra intencional puede llegar a ser subjetiva.

Observa las siguientes imágenes. En ellas se explica un ejemplo de cómo elegir una muestra. Aunque cada caso es diferente, este ejemplo puede ayudarte a hacerte una idea de las cosas que hay que tener en cuenta a la hora de elegir una buena muestra.



Por ejemplo...



## ¿Cómo se elige una muestra?

Queremos saber cuales serían las propuestas que haría un andaluz o andaluza para mejorar la situación medioambiental en Andalucía.

Elegimos una **MUESTRA** entre la población andaluza

### 1º ¿Qué entendemos cómo población andaluza?

Nacidos/as en Andalucía

Censados/as en Andalucía

Cualquier persona que  
habe en este momento  
en una localidad andaluza

... y elegimos ...

Cualquier persona que  
habe en este momento  
en una localidad andaluza

### 2º ¿Qué número de personas formarán parte de la muestra?

Cuántas más, mejor.

Para facilitar el estudio decidimos formar  
una muestra de **100** personas

### 3º ¿Cuál será la configuración de la muestra?

Habrà que tener en cuenta:

Sexo

Edad

Estatus social

Localidad

Etc.....

<b>Sexo:</b> Hay que encuestar tanto a hombres como a mujeres en proporciones cercanas al 50%	<b>Estatus social:</b> Procurar que estén representados los distintos grupos de la población, diferenciando, por ejemplo: El nivel de estudios, la ocupación laboral, .....
<b>Edad:</b> Deberían de estar representados todos los grupos de edades (respetando la proporción poblacional) <ul style="list-style-type: none"><li>-Mayores de 65 años</li><li>-Entre 40 y 65 años</li><li>-Entre 25 y 40 años</li><li>-Menores de 25 años</li></ul>	<b>Localidad:</b> Se debe tener en cuenta el lugar en el que habita el entrevistado/a, por ejemplo, determinando si es una población de pocos habitantes (rural), o una ciudad.  También habría que tener en cuenta que hubiese representantes de las ocho provincias andaluzas

**4º ¿Cómo elegimos el entrevistado/a?**

se debe elegir al entrevistado/a de forma ALEATORIA (al azar).

El método más usado es la llamada telefónica, o la entrevista a pie de calle.

Por ejemplo imaginemos que nos queda por completar la muestra:

MUJER  
MAYOR DE 65 AÑOS  
CON ESTUDIOS SUPERIORES  
HABITANTE DE UNA LOCALIDAD DE MENOS DE 3000 hab.  
DE LA PROVINCIA DE CÁDIZ

Sólo podremos entrevistar a una persona que cumpla todas estas características ... así que ¡a por ella!

Es recomendable ir planificando la composición de la muestra, para evitar que al final nos queden casos tan complicados de encontrar como el ejemplo aquí expuesto.

En el apartado de recursos del tema puedes ver el ejemplo anterior en una presentación llamada "**La elección de la muestra**".

Aunque, por supuesto, la mayoría de las veces el **sentido común** nos dirá cuál es la mejor manera de elegir la muestra. Prueba tú con esta... con estos ejercicios.

### Comprueba que lo has entendido

4. Estás realizando un estudio estadístico para conocer la satisfacción de la gente del barrio con el nuevo polideportivo. ¿Qué forma de elegir la muestra crees que es mejor?

- Preguntar a 50 personas que estén en el polideportivo.
- Preguntar a 50 personas de tus amistades.
- Elegir al azar 50 números de teléfono de casas del barrio, llamar y preguntar.
- Preguntar a 50 personas que estén por la mañana comprando en el mercado.





## Comprueba que lo has entendido

5. Se quiere estudiar el nivel de contaminación del agua de un río. Escoge la opción más adecuada para elegir la muestra:

- a. Se cogería una muestra de agua al azar de cualquier zona del cauce del río.
- b. Se tomarían varias muestras de agua al azar de distintas zonas a lo largo del cauce del río y en distintos periodos de tiempo.
- c. Se tomaría una muestra de agua al lado de una fábrica que vierte sus residuos directamente al cauce del río.
- d. Se tomaría una muestra de agua en el lugar de nacimiento del río.



## Preparamos los datos para analizarlos

Ya tenemos dado el primer paso de nuestro estudio estadístico. Hemos decidido qué queremos saber, de quién lo queremos saber y el modo en el que vamos a recoger los datos que necesitaremos para saberlo.

Ahora tendremos que realizar las encuestas que hemos diseñado o tomar las medidas que hemos previsto.

El resultado de este trabajo será un montón de datos que, si no los organizamos convenientemente, no nos ayudarán a resolver nuestro problema, que no es otro que el de obtener una información "fiable" y "objetiva" sobre un determinado tema.

Disponemos de dos formas de organizar los datos de modo que "les podamos sacar jugo":

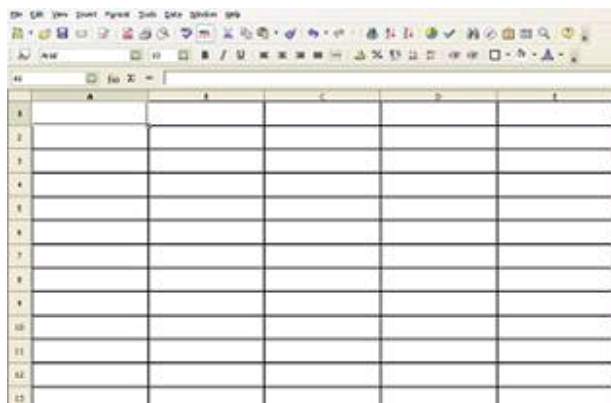
- Ordenarlos en **tablas**.
- Representarlos en **gráficas**.

Lo normal es hacer las dos cosas y... eso es lo que vas a aprender ahora.



¡Es fácil y hasta entretenido!

## Organizamos los datos en una tabla



Organizar los datos en tablas es la opción más extendida y eficaz, sólo basta abrir cualquier modelo de hoja de cálculo o base de datos en nuestra pantalla de ordenador y observar que su formato es por defecto una tabla.

Es muy importante **la organización de los datos en forma de tabla**, ya que **los hace más comprensibles y facilita los cálculos**.

Observa en los dos ejemplos siguientes cómo se organizan los datos en una tabla, tanto para una variable cuantitativa como para una cualitativa.



Por ejemplo...

## Formación de una tabla de datos

Se ha realizado una encuesta a 50 personas preguntando por el número de parajes o parques naturales que conoce de Andalucía, siendo sus respuestas:

0	1	2	1	0	3	4	4	3
1	1	2	2	2	5	3	2	1
3	3	3	2	2	3	4	0	2
3	5	2	1	0	1	2	1	0
4	2	2	3	0	0	1	2	3
2	2	3	2	1				

Se puede observar que **hay sólo seis datos diferentes** que son: 0, 1, 2, 3, 4, 5

0	1	2	1	0	3	4	4	3
1	1	2	2	2	5	3	2	1
3	3	3	2	2	3	4	0	2
3	5	2	1	0	1	2	1	0
4	2	2	3	0	0	1	2	3
2	2	3	2	1				

Al número de veces que se repite cada dato se le llama **frecuencia absoluta** del dato

Agrupamos los datos en una tabla de frecuencias absolutas:

Dato ( $x_i$ )	Frec. Abs. ( $f_i$ )
0	7
1	
2	
3	
4	
5	

<del>0</del>	1	2	1	<del>0</del>	3	4	4	3
1	1	2	2	2	5	3	2	1
3	3	3	2	2	3	4	<del>0</del>	2
3	5	2	1	<del>0</del>	1	2	1	<del>0</del>
4	2	2	3	<del>0</del>	<del>0</del>	1	2	3
2	2	3	2	1				

<del>0</del>	<del>1</del>	<del>2</del>	<del>1</del>	<del>0</del>	<del>3</del>	<del>4</del>	<del>4</del>	<del>3</del>
<del>1</del>	<del>1</del>	<del>2</del>	<del>2</del>	<del>2</del>	<del>5</del>	<del>3</del>	<del>2</del>	<del>1</del>
<del>3</del>	<del>3</del>	<del>3</del>	<del>2</del>	<del>2</del>	<del>3</del>	<del>4</del>	<del>0</del>	<del>2</del>
<del>3</del>	<del>5</del>	<del>2</del>	<del>1</del>	<del>0</del>	<del>1</del>	<del>2</del>	<del>1</del>	<del>0</del>
<del>4</del>	<del>2</del>	<del>2</del>	<del>3</del>	<del>0</del>	<del>0</del>	<del>1</del>	<del>2</del>	<del>3</del>
<del>2</del>	<del>2</del>	<del>3</del>	<del>2</del>	<del>1</del>				

Dato ( $x_i$ )	Frec. Abs. ( $f_i$ )
0	7
1	10
2	16
3	11
4	4
5	2
<b>Total</b>	<b>50</b>

Por ejemplo...



### Formación de una tabla de datos

Se ha realizado un estudio estadístico preguntando a 50 personas sobre las posibles causas del avance de la desertificación en la zona mediterránea española. La encuesta constaba de las siguientes opciones:



☐ Condiciones climáticas semiáridas que afectan a grandes zonas, sequías estacionales, extrema variabilidad de las lluvias y lluvias súbitas de gran intensidad.



☐ Suelos pobres con marcada tendencia a la erosión, propensos a la formación de cortezas superficiales.



☐ Un relieve desigual, con laderas escarpadas y paisajes muy diversificados.



☐ Grandes pérdidas de la cubierta forestal a causa de repetidos incendios de bosques.



☐ Condiciones de crisis en la agricultura tradicional, con el consiguiente abandono de tierras y deterioro del suelo y de las estructuras de conservación del agua.



☐ Ocasional explotación insostenible de los recursos hídricos, que es causa de graves daños ambientales, incluidos la contaminación química, la salinización y el agotamiento de los acuíferos.










☐ Concentración de la actividad económica en las zonas costeras como resultado del crecimiento urbano, las actividades industriales, el turismo y la agricultura de regadío.

Cada persona encuestada marcaba con una **X** la respuesta que creía el principal motivo de desertificación:



☒ Grandes pérdidas de la cubierta forestal a causa de repetidos incendios de bosques.

	Dato ( $x_i$ )	Frec. Abs. ( $f_i$ )
	Condiciones climáticas semiáridas que afectan a grandes zonas, sequías estacionales, extrema variabilidad de las lluvias y lluvias súbitas de gran intensidad.	8
	Suelos pobres con marcada tendencia a la erosión, propensos a la formación de cortezas superficiales.	6
	Un relieve desigual, con laderas escarpadas y paisajes muy diversificados.	5
	Grandes pérdidas de la cubierta forestal a causa de repetidos incendios de bosques.	12
	Condiciones de crisis en la agricultura tradicional, con el consiguiente abandono de tierras y deterioro del suelo y de las estructuras de conservación del agua.	7
	Ocasional explotación insostenible de los recursos hídricos, que es causa de graves daños ambientales, incluidos la contaminación química, la salinización y el agotamiento de los acuíferos.	10
	Concentración de la actividad económica en las zonas costeras como resultado del crecimiento urbano, las actividades industriales, el turismo y la agricultura de regadío.	2

**Total: 50**



En el apartado de recursos del tema puedes ver los ejemplos anteriores en dos presentaciones llamadas “*Tabla para una variable cuantitativa*” y “*Tabla para una variable cualitativa*”.

## Comprueba que lo has entendido

6. Anotamos el nombre de las distintas especies animales que se encuentran en el parque y resulta la siguiente variable cualitativa:

*paloma, gorrión, gato, perro, koi, ardilla, hormiga, mosquito, mosca, araña, cigüeña, goldfish, mirlo, avispa, rana, lagartija, salamandresa, gusano, ratón, topo, urraca, golondrina, libélula, carpa, niños, grillo, escarabajo, cochinilla, pato, cisne.*

¿Cuál de las siguientes tablas de frecuencias es la correcta?

Clase de animal	Frecuencia
Aves	8
Peces	1
Mamíferos	7
Reptiles	3
Anfibios	2
Invertebrados	9
Tamaño de la muestra	30

OPCIÓN A



Clase de animal	Frecuencia
Aves	8
Peces	3
Mamíferos	6
Reptiles	2
Anfibios	1
Invertebrados	10
Tamaño de la muestra	30

OPCIÓN B

## Representamos los datos en una gráfica

Una **gráfica estadística** es la **mejor forma de presentar** toda la **información que se ha recogido**. Con una simple "ojeada" nos permite distinguir, sin dificultad alguna, que opción es la preferida por los encuestados.

La prensa diaria está llena de ejemplos (prueba y verás como en cualquier diario que tengas a mano aparece como mínimo una gráfica estadística).

Existen muchos modelos de gráficas estadísticas, aunque los más difundidos son la **gráfica de barras** y el **diagrama de sectores**:

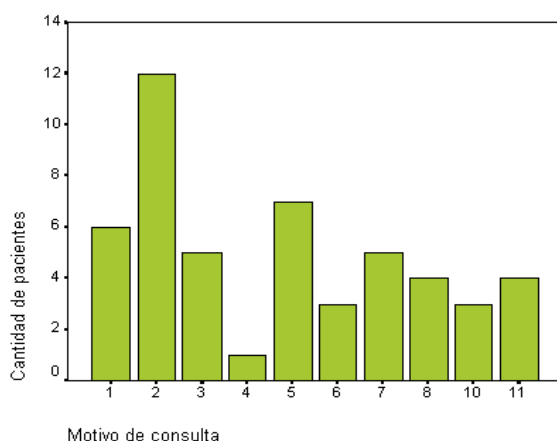


DIAGRAMA DE BARRAS

### DEPORTE MÁS PRACTICADO

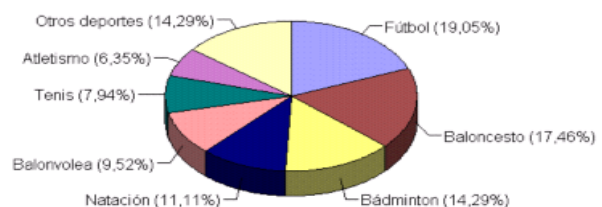


GRÁFICO DE SECTORES

En los siguientes ejemplos puedes ver con detalle cómo se construyen con ayuda de la informática (que para eso está, para ayudarnos en estos y otros menesteres)

Por ejemplo...



## Creación de una gráfica de barras



Se ha realizado una encuesta a 50 personas preguntando por el número de parajes o parques naturales que conoce de Andalucía, siendo sus respuestas:

Dato ( $x_i$ )	Frec. Abs. ( $f_i$ )
0	7
1	10
2	16
3	11
4	4
5	2
<b>Total</b>	<b>50</b>



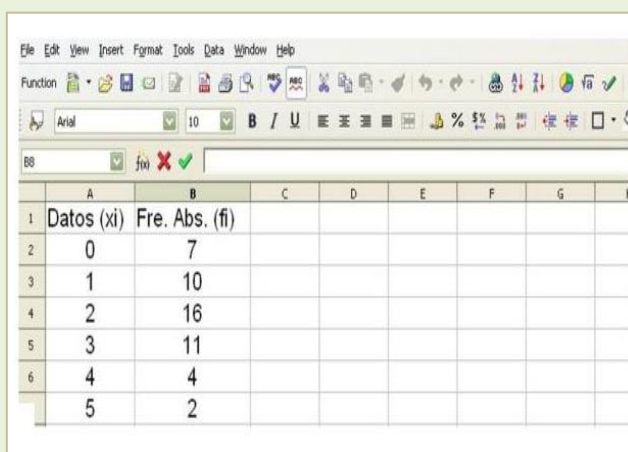
Con estos datos vamos a crear un **gráfico de barras**.

Para ello usamos una **hoja de cálculo** (en este caso OpenOffice Calc)

Usamos una hoja de cálculo (tipo OpenOffice Calc o Excel)

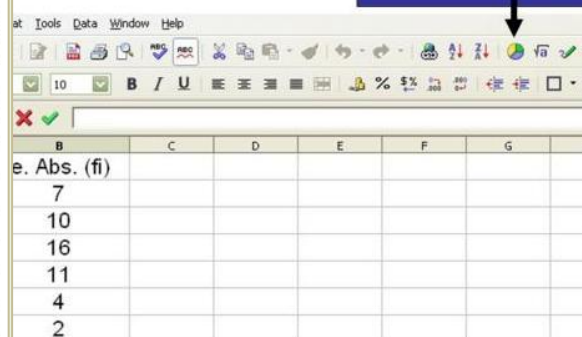
### Paso 1

Introducimos la tabla de datos, asociados a sus respectivas frecuencias absolutas, en una hoja de cálculo.



	A	B	C	D	E	F	G
1	Datos ( $x_i$ )	Frec. Abs. ( $f_i$ )					
2	0	7					
3	1	10					
4	2	16					
5	3	11					
6	4	4					
7	5	2					

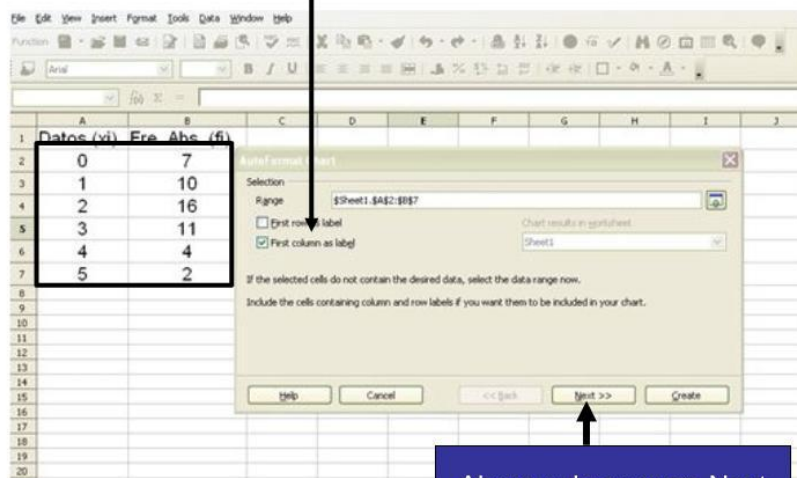
Pulsamos sobre el botón de creación de gráficos



	B	C	D	E	F	G
1	e. Abs. ( $f_i$ )					
2	7					
3	10					
4	16					
5	11					
6	4					
7	2					

### Paso 2

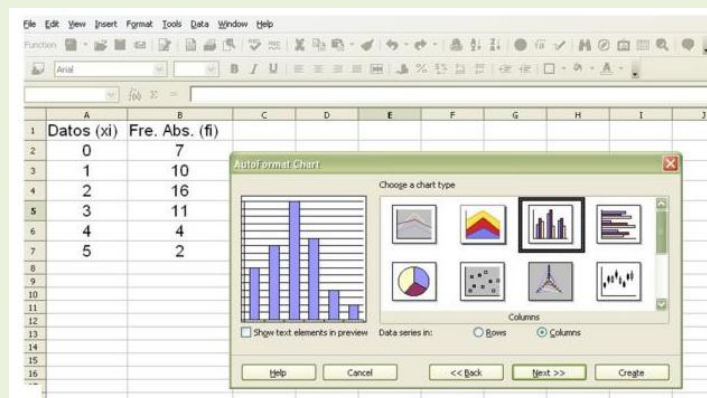
Marcamos el área en la que queremos que se dibuje la gráfica y su tamaño. Después señalamos las dos columnas en las que se encuentran datos y frecuencias, marcando la opción **First column as label** (en la primera columna están los datos).



Ahora pulsamos en Next

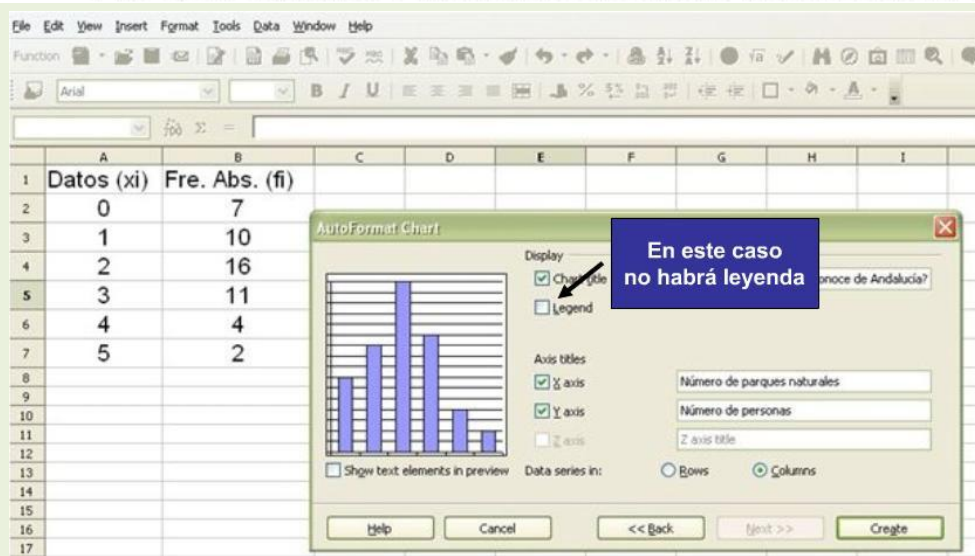
### Paso 3

Elegimos el tipo gráfico de barras en Columnas y pulsamos Next



### Paso 4

Rellenamos las etiquetas, con el título del gráfico, del eje de abscisas X (datos) y del eje de ordenadas Y (frecuencias absolutas). Pulsamos el botón **Create**.

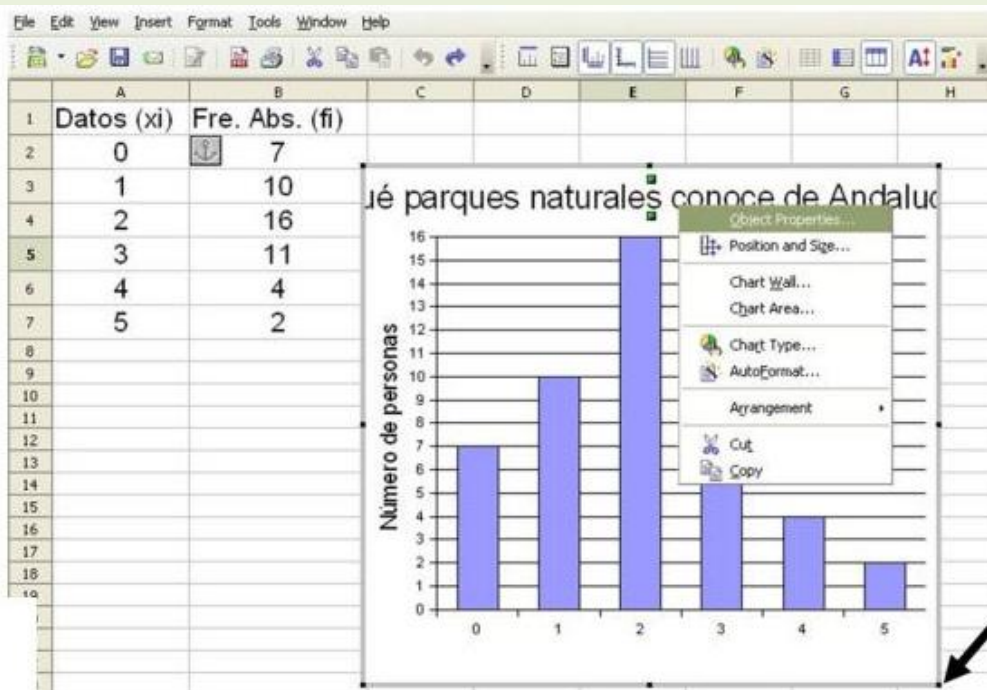


En este caso  
no habrá leyenda



### Paso 5

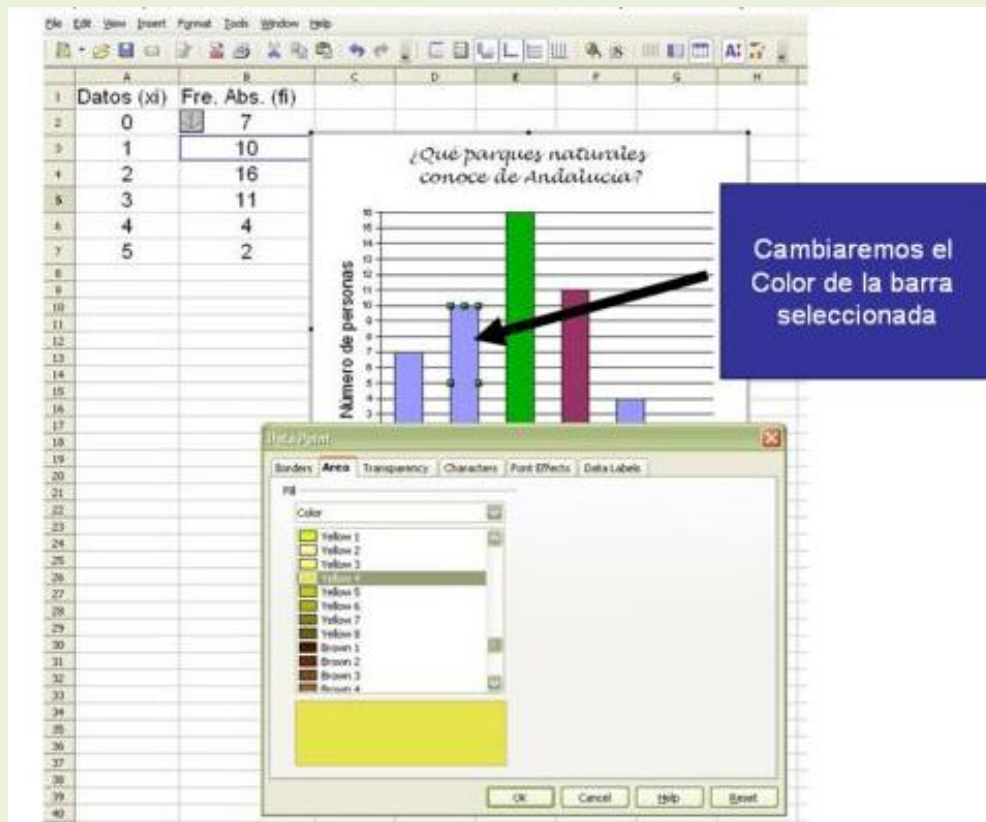
Una vez creado el gráfico pulsamos sobre cualquier elemento que queramos cambiar desplegándose un menú y seleccionando la opción **Object Properties**



También podemos cambiar el tamaño del gráfico pulsando en los bordes

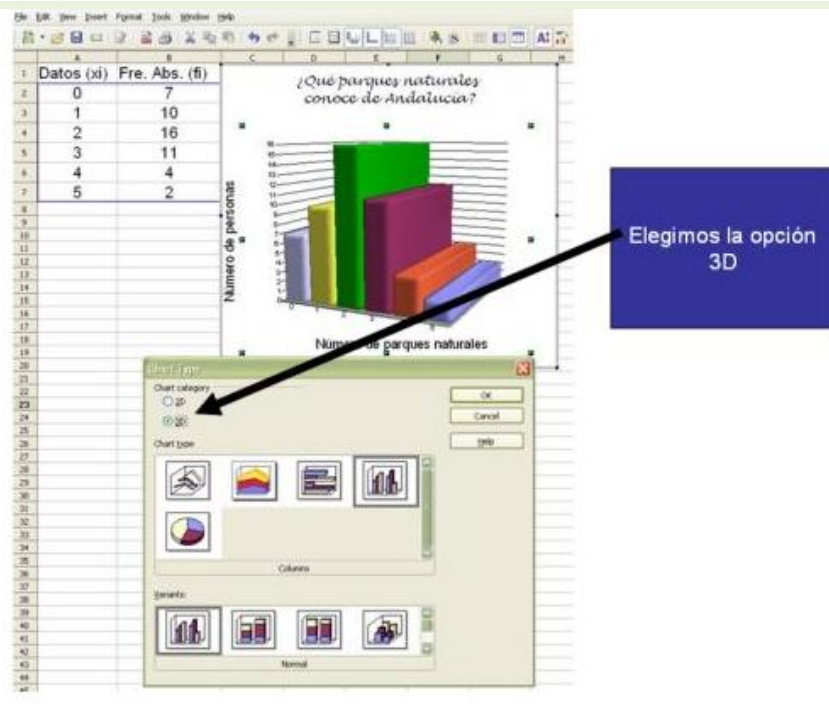
### Paso 6

Si pulsamos sobre cualquier barra podemos cambiar su color, ponerle borde, hacer que aparezca su valor en número o en porcentaje, etc.

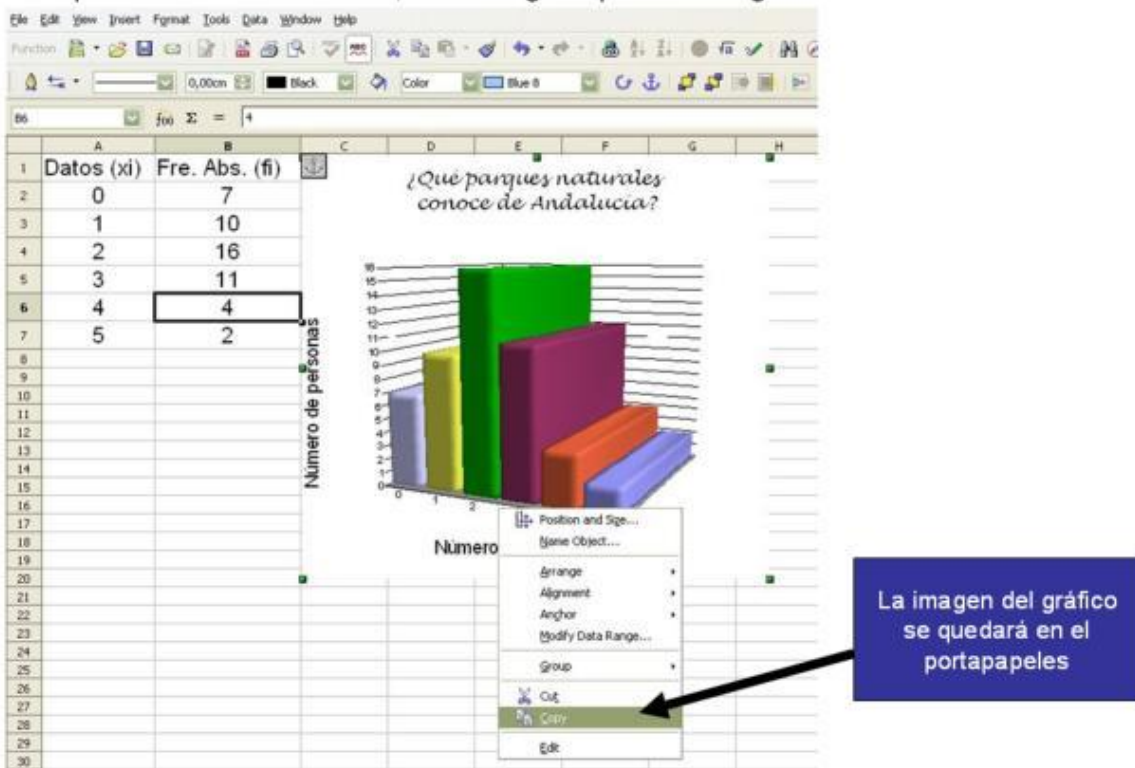


## Paso 7

Podemos cambiar la vista a 3D usando la opción **Chart Type**



Podemos copiar el cuadro de imagen en el que está el gráfico y pegarlo en cualquier documento de texto, se trata igual que una imagen



**Paso 8**

Aquí tienes la copia final de nuestro gráfico de barras.



Esperamos que esta guía te haya sido útil

Por ejemplo...



## Creación de un diagrama de sectores

Se ha realizado un estudio estadístico preguntando a 50 personas sobre las posibles causas del avance de la **desertificación** en la zona mediterránea española. La encuesta constaba de las siguientes opciones:



*Condiciones climáticas semiáridas que afectan a grandes zonas, sequías estacionales, extrema variabilidad de las lluvias y lluvias súbitas de gran intensidad*



*Suelos pobres con marcada tendencia a la erosión, propensos a la formación de cortezas superficiales.*



*Un relieve desigual, con laderas escarpadas y paisajes muy diversificados.*



*Grandes pérdidas de la cubierta forestal a causa de repetidos incendios de bosques.*



*Condiciones de crisis en la agricultura tradicional, con el consiguiente abandono de tierras y deterioro del suelo y de las estructuras de conservación del agua.*









*Ocasional explotación insostenible de los recursos hídricos, que es causa de graves daños ambientales, incluidos la contaminación química, la salinización y el agotamiento de los acuíferos.*



*Concentración de la actividad económica en las zonas costeras como resultado del crecimiento urbano, las actividades industriales, el turismo y la agricultura de regadío.*

El número de personas encuestadas que eligieron cada opción fue:



 Condiciones climáticas semiáridas que afectan a grandes zonas, sequías estacionales, extrema variabilidad de las lluvias y lluvias súbitas de gran intensidad <b>8</b>	 Suelos pobres con marcada tendencia a la erosión, propensos a la formación de cortezas superficiales. <b>6</b>
 Un relieve desigual, con laderas escarpadas y paisajes muy diversificados. <b>5</b>	 Grandes pérdidas de la cubierta forestal a causa de repetidos incendios de bosques. <b>12</b>
 Condiciones de crisis en la agricultura tradicional, con el consiguiente abandono de tierras y deterioro del suelo y de las estructuras de conservación del agua. <b>7</b>	 Ocasional explotación insostenible de los recursos hídricos, que es causa de graves daños ambientales, incluidos la contaminación química, la salinización y el agotamiento de los acuíferos. <b>10</b>
 Concentración de la actividad económica en las zonas costeras como resultado del crecimiento urbano, las actividades industriales, el turismo y la agricultura de regadío. <b>2</b>	

Para facilitar el trabajo resumiremos cada una de las opciones como:

 <b>meteorología</b> <b>8</b>	 <b>tipo de suelo</b> <b>6</b>	 <b>relieve</b> <b>5</b>
 <b>crisis agrícola</b> <b>7</b>	 <b>incendios</b> <b>12</b>	 <b>sequía</b> <b>10</b>
 <b>desarrollo de la costa</b> <b>2</b>		

Usamos una hoja de cálculo (tipo OpenOffice Cal o Excel)

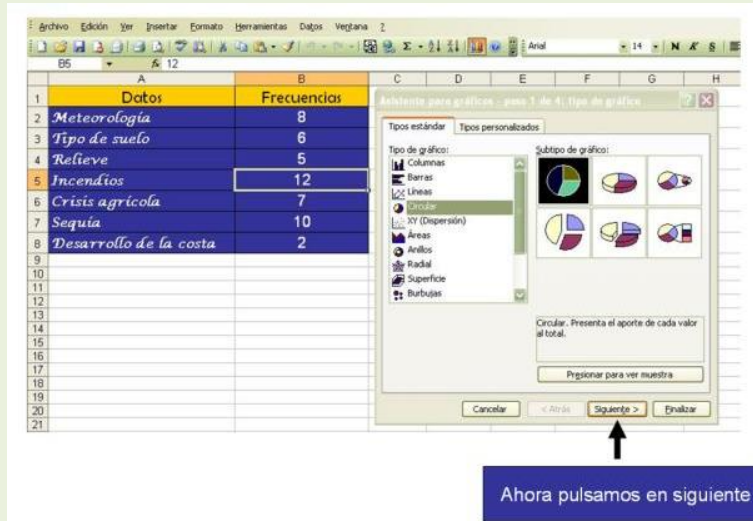
**Paso 1** Introducimos la tabla de datos, asociados a sus respectivas frecuencias absolutas, en una hoja de cálculo.

Pulsamos sobre el botón de creación de gráficos



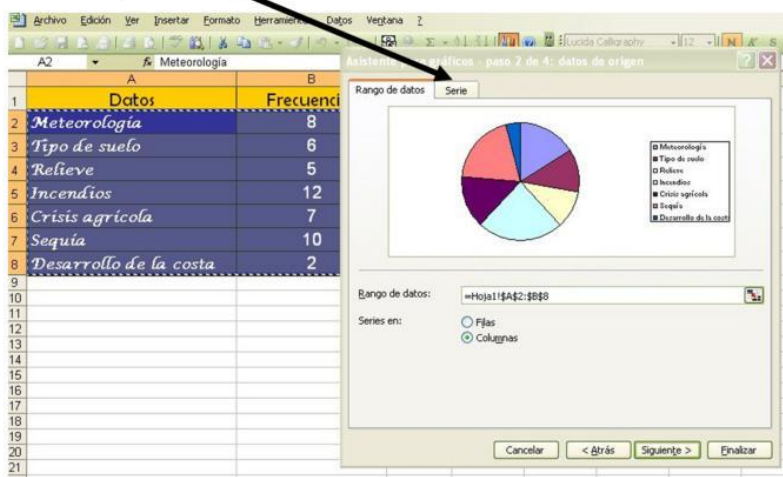
	A	B	C	D	E	F	G
1	<b>Datos</b>	<b>Frecuencias</b>					
2	Meteorología	8					
3	Tipo de suelo	6					
4	Relieve	5					
5	Incendios	12					
6	Crisis agrícola	7					
7	Sequía	10					
8	Desarrollo de la costa	2					
9							

**Paso 2** Se abre una pantalla en la que seleccionamos el tipo de gráfico (en este caso circular)

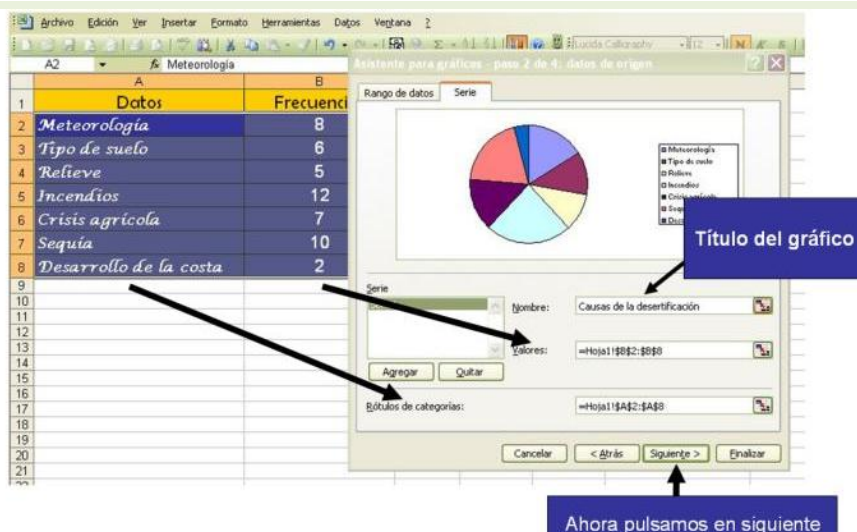


Ahora pulsamos en siguiente

**Paso 3** Marcamos las dos columnas con los datos y la frecuencias (rango de datos) y pulsamos en la pestaña **Serie**



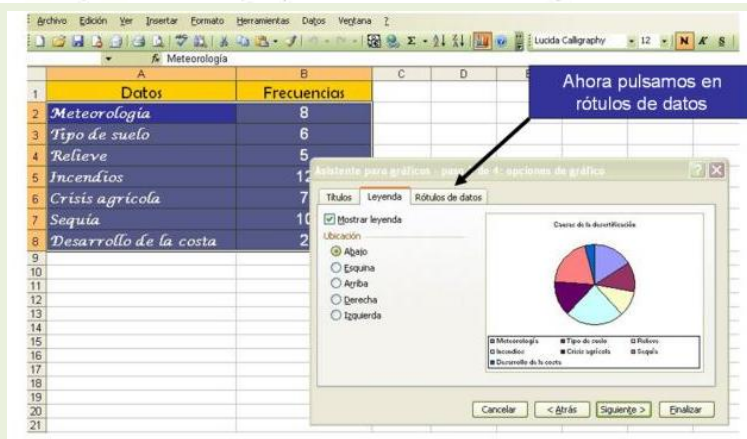
**Paso 4** Le damos un nombre al gráfico, en este caso "Causas de la desertificación"



Ahora pulsamos en siguiente

### Paso 5

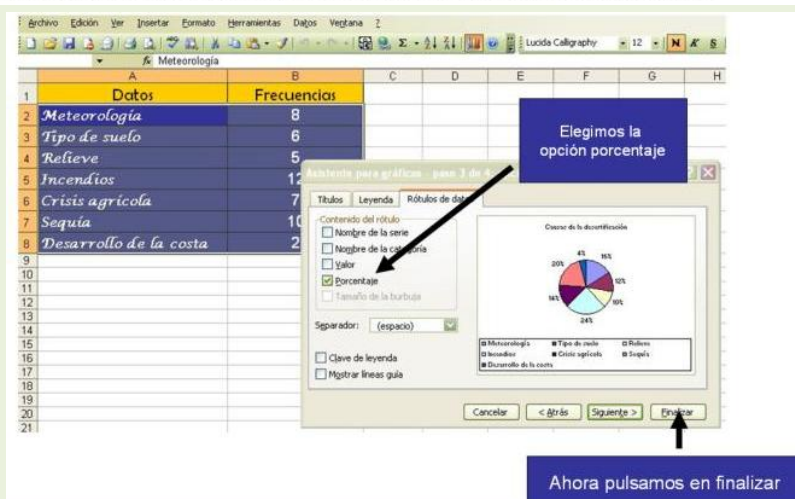
En la opción **Leyenda** podemos marcar desde que no aparezca, hasta su posición en la gráfica. En el ejemplo la hemos situado **abajo**.



Ahora pulsamos en **rótulos de datos**

### Paso 6

Esta pestaña nos permite mostrar el valor de cada sector del diagrama, su porcentaje equivalente, o el nombre de la categoría a la que corresponde.



Elegimos la opción **porcentaje**

Ahora pulsamos en **finalizar**

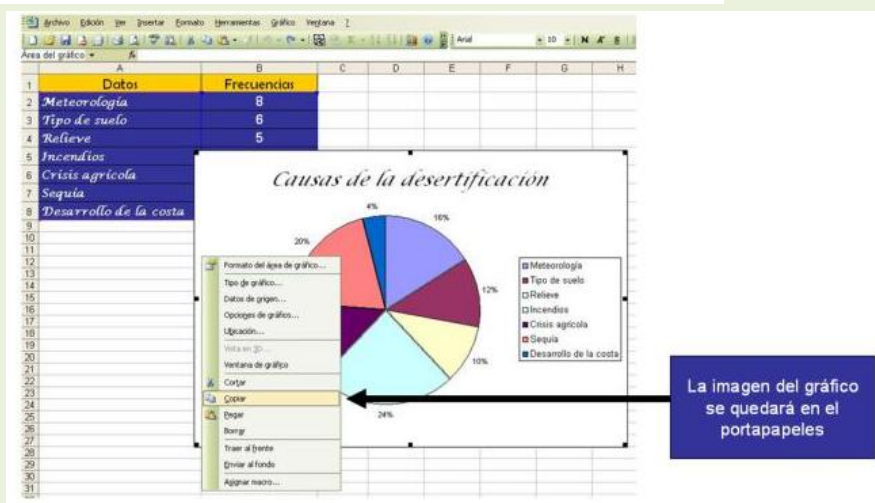
### Paso 7

Ahora podemos seleccionar cualquier zona del gráfico y cambiar el tamaño, tipo de letra, ubicación de la leyenda, colores, etc.

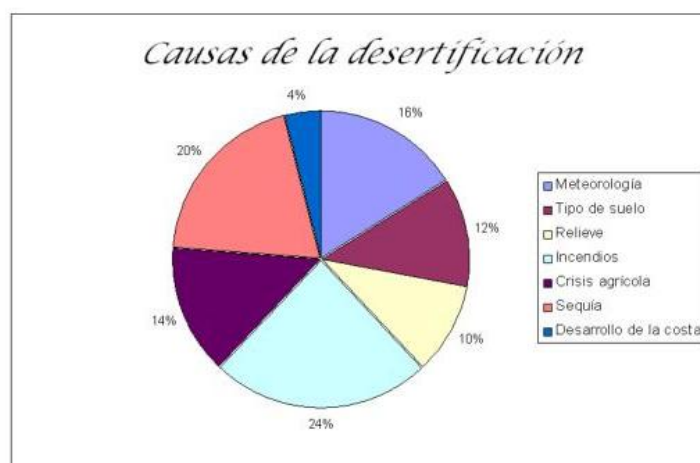




Podemos copiar el cuadro de imagen en el que está el gráfico y pegarlo en cualquier documento de texto, se trata igual que una imagen



Aquí tienes la copia final de nuestro diagrama de sectores.



Esperamos que esta guía te haya sido útil

En el apartado de recursos del tema puedes ver los ejemplos anteriores en dos presentaciones llamadas “*Creación de un diagrama de barras*” y “*Creación de un gráfico de sectores*”.

Comprueba que lo has entendido

7. Vamos a ver qué tal se te da asociar una tabla de datos con la gráfica que los representa. ¿Cuál de los gráficos de abajo se corresponde con la tabla de datos que ves a continuación?

Residuos urbanos recogidos clasificados en Andalucía durante el año 2006.	Toneladas
Aceites minerales usados (no incluye los de tipo alimentario)	14650
Medicamentos no utilizados	610
Residuos sanitarios y biológicos	136640
Residuos metálicos	29934
Vidrio	54861
Papel y cartón	82252
Caucho (neumáticos.)	5209
Plásticos (excepto embalajes)	7521
Madera	6635
Ropa y residuos textiles	803



## Le sacamos jugo a los datos



Ya estás muy cerca de terminar tu estudio estadístico. Con los datos bien ordenaditos esto ya parece otra cosa ¿verdad?

Pero ahora **llega la hora de las cuentas**. Para poder "sacar jugo" a tanto dato es necesario resumir de alguna forma toda la información que contienen. Para esto, los matemáticos calculan unos numeritos: los **parámetros estadísticos**.

Hay un montón de parámetros estadísticos... Pero no te preocupes, tú solo tendrás que manejar unos cuantos de ellos, los más sencillos. Alguno ya lo manejas desde hace tiempo, probablemente sin saber que se trata de un parámetro estadístico. Nos referimos a la media aritmética ¿O no has calculado más de una vez la nota media que has sacado en una asignatura...?

En este tema vas a estudiar varios parámetros estadísticos (concretamente seis). Pero los podemos agrupar en solo dos tipos:



$\bar{x}$   $\mu$   
 $\sigma$   $\Sigma$

TIPO DE PARÁMETRO	PARÁMETROS ESTADÍSTICOS
CENTRALES	<ul style="list-style-type: none"> <li>▶ MEDIA ARITMÉTICA</li> <li>▶ MODA</li> </ul>
DE DISPERSIÓN	<ul style="list-style-type: none"> <li>▶ RANGO</li> <li>▶ VARIANZA</li> <li>▶ DESVIACIÓN TÍPICA</li> <li>▶ COEFICIENTE DE VARIACIÓN</li> </ul>



Los **parámetros estadísticos** son...  
 una forma de **resumir la información en un único valor numérico**.

### Parámetros centrales



Los parámetros centrales son números que...  
**representan de forma global al conjunto de los datos.**

El parámetro central más empleado es la **media aritmética** (a la que normalmente todo el mundo se refiere simplemente como "la media"). Se suele representar con los símbolos  $\bar{x}$  o  $\mu$  y es muy fácil de calcular.



**Se suma el valor de todos los datos y...**  
**se divide entre el tamaño de la muestra (el número total de datos)**





### Por ejemplo...



Supón que en una calle hay tres casas con jardín. En uno de los jardines hay 7 maceteros, en otro 3 y en el tercero hay 2. Entonces, el número medio de maceteros que hay en los jardines de esas tres casas (la media aritmética) será:

$$\bar{x} = \frac{7 + 3 + 2}{3} = 4$$

Si quisiéramos tener el mismo número de maceteros en los tres jardines, pero que el número total de maceteros no cambiase, tendríamos que poner 4 maceteros en cada jardín.

### Otro ejemplo...

Supón que las calificaciones de todas las tareas que vas a hacer en este trimestre son:

6, 7, 6, 9, 8, 6, 5, 4, 6, 7, 6, 5, 5, 6.

Entonces, la nota que sacarás en el apartado de tareas, que será la media aritmética de esas calificaciones se calculará...

$$\bar{x} = \frac{6 + 7 + 6 + 9 + 8 + 6 + 5 + 4 + 6 + 7 + 6 + 5 + 5 + 6}{14} = 6,14$$

Es decir, sumando todas las calificaciones y dividiendo entre el número total de calificaciones (14 en este caso)



Fíjate en que la puntuación total obtenida sumando las 14 calificaciones es 86. Si en cada uno de las 14 tareas hubieses sacado un 6,14 también obtendrías una puntuación total de 86.

Observa que hay algunas calificaciones que se repiten ¿verdad? Podríamos haber hecho la misma cuenta de un modo más sencillo; así:

$$\bar{x} = \frac{6 \cdot 6 + 7 \cdot 2 + 9 \cdot 1 + 8 \cdot 1 + 5 \cdot 3 + 4 \cdot 1}{14} = 6,14$$

Es decir, multiplicando cada calificación por el número de veces que aparece (su frecuencia absoluta), sumando después todos los resultados y dividiendo al final entre el número total de calificaciones (que siguen siendo 14, por supuesto)

Aquí puedes ver otro ejemplo de cálculo de la media aritmética...

### Por ejemplo...



### Calculamos la media aritmética

Se ha realizado una encuesta a 50 personas preguntando por el número de parajes o parques naturales que conoce de Andalucía, siendo sus respuestas:

Dato ( $x_i$ )	Frec. Abs. ( $f_i$ )
0	7
1	10
2	16
3	11
4	4
5	2
<b>Total</b>	<b>50</b>



Vamos a calcular la media aritmética. Para ello debemos SUMAR todos los datos y DIVIDIR el resultado entre el número de personas encuestadas. Podemos hacerlo de tres formas diferentes:

1ª. Sumamos directamente los datos uno a uno

0 + 1 + 2 + 1 + 0 + 3 + 4 + 4 + 3 +  
 1 + 1 + 2 + 2 + 2 + 5 + 3 + 2 + 1 +  
 3 + 3 + 3 + 2 + 2 + 3 + 4 + 0 + 2 +  
 3 + 5 + 2 + 1 + 0 + 1 + 2 + 1 + 0 +  
 4 + 2 + 2 + 3 + 0 + 0 + 1 + 2 + 3 +  
 2 + 2 + 3 + 2 + 1 = 101

Se divide entre el número de entrevistas:  $\frac{101}{50} = 2.02$

Por tanto el número medio de parques o parajes naturales que conoce un andaluz será de 2.02 (que podemos aproximar a 2)

2ª. Sumamos con la ayuda de la tabla de frecuencias

Dato ( $x_i$ )	Frec. Abs. ( $f_i$ )	$x_i \cdot f_i$
0	7	0
1	10	10
2	16	32
3	11	33
4	4	16
5	2	10
	<b>50</b>	<b>101</b>

Para completar esta nueva columna multiplicamos cada dato por su frecuencia absoluta

Ejemplo:  $2 \cdot 16 = 32$

Sumamos toda la columna y el resultado es la suma total de datos

Se divide entre el número de entrevistas:  $\frac{101}{50} = 2.02$

Por tanto el número medio de parques o parajes naturales que conoce un andaluz será de 2.02 (que podemos aproximar a 2)

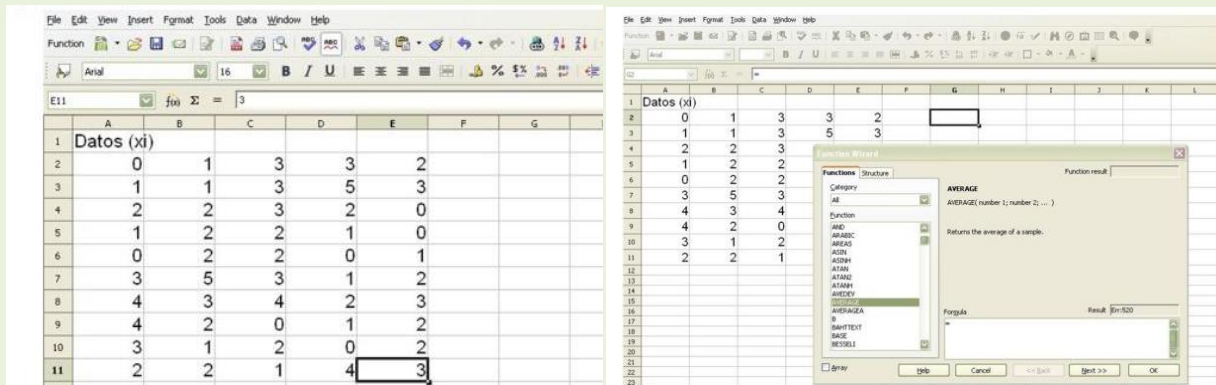
3ª. Usamos una hoja de cálculo (tipo Open Office Cal o Excel)

Paso 1

Introducimos los datos uno a uno en una hoja de cálculo, cada dato ocupará una casilla

Paso 2

Elegimos una casilla en blanco y elegimos la acción INSERT y después FUNCTION. De la lista de funciones, elegimos AVERAGE...

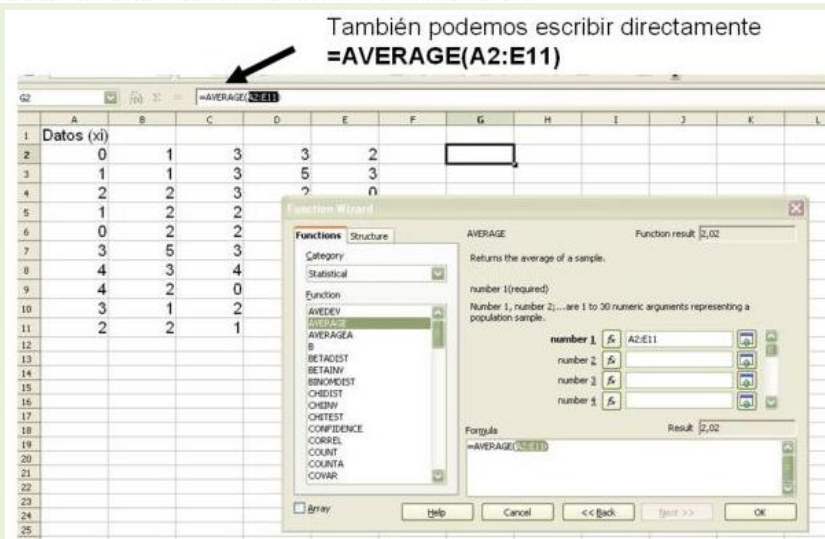


The image shows two screenshots of a spreadsheet application. The left screenshot shows a grid with data entered in columns A through E. The right screenshot shows the 'Function Wizard' dialog box with 'AVERAGE' selected, and the 'Result' field showing the calculated average of the selected data range.

### Paso 3

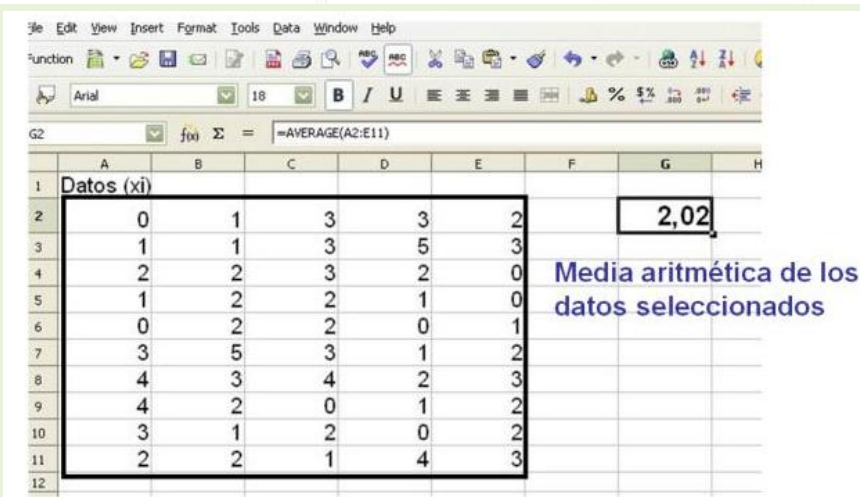
Pulsamos el botón NEXT y seleccionamos con el ratón todas las casillas en las que estén escritos los datos, después aceptamos

También podemos escribir directamente  
**=AVERAGE(A2:E11)**



### Paso 4

En la casilla seleccionada aparecerá el valor de la media aritmética



Media aritmética de los datos seleccionados

La fórmula general para calcular la media aritmética es...

(no te asustes, que no muerde)

$$\bar{x} = \frac{\sum_{i=1}^n x_i \cdot f_i}{N}$$

Sumatoria

Del producto de los datos por su frecuencia

Media aritmética

Dividido

Tamaño de la muestra (nº total de datos)

En el apartado de recursos del tema puedes ver este ejemplo en una presentación llamada *“Así se calcula de la media”*.



## Mucho cuidado...

La media aritmética está muy bien. Es muy utilizada porque es fácil de calcular y, por lo general, bastante representativa. Pero también tiene sus pegas:

- Sólo puede calcularse si la variable que estamos estudiando es cuantitativa.
- Si hay datos muy "extremos" su significado pierde mucho sentido.

Por ejemplo, si de 100 jóvenes hay 50 que salen 4 fines de semana al mes y otros 50 que no salen nunca (salen 0 veces), la media aritmética nos dice que los jóvenes salen una media de 2 fines de semana al mes. Pero este dato no es significativo, puesto que la mitad de los jóvenes (pobreticos ellos) no salen nunca. .



No es ésta la moda a la que nos referimos.

En casos en los que la media no se puede calcular o no es lo bastante representativa, se emplea la **moda**.

Se suele representar con el símbolo  $M_0$  y es aún más fácil de calcular de la moda porque... ¡no hace falta hacer cuentas! (al menos en los casos más sencillos).



La moda es **el dato que más se repite**, es decir, el que tiene mayor frecuencia absoluta.

## Por ejemplo...



En el ejemplo de las calificaciones (el que hay un poco más arriba), la moda es 6, puesto que es la calificación de mayor frecuencia absoluta, la calificación que más se repite.

En el ejemplo de los maceteros... la moda es un parámetro que no tiene sentido, puesto que todos los datos se repiten el mismo número de veces: cada dato aparece una sola vez.

Como ves en los ejemplos, la moda no siempre nos ofrece información útil. También **puede pasar que en una serie de datos haya varias modas**, es decir, varios datos con la misma frecuencia absoluta y que sea la mayor de todas.

Observa, en las siguientes imágenes, cómo de fácil es obtener la moda:

## Por ejemplo...



## Calculamos la moda

Para una variable cuantitativa:

Se ha realizado una encuesta a 50 personas preguntando por el número de parajes o parques naturales que conoce de Andalucía, siendo sus respuestas:

Dato ( $x_i$ )	Frec. Abs. ( $f_i$ )
0	7
1	10
2	16
3	11
4	4
5	2
<b>Total</b>	<b>50</b>








La moda es el dato más frecuente.

En este caso la Moda es

"Conocer 2 parajes o parques naturales de Andalucía"

### Para una variable cualitativa:

Se ha realizado un estudio estadístico preguntando a 50 personas sobre las posibles causas del avance de la **desertificación** en la zona mediterránea española. La encuesta constaba de las siguientes opciones:

	Dato ( $x_i$ )	Frec. Abs. ( $f_i$ )
	Condiciones climáticas semiáridas que afectan a grandes zonas, sequías estacionales, extrema variabilidad de las lluvias y lluvias súbitas de gran intensidad	8
	Suelos pobres con marcada tendencia a la erosión, propensos a la formación de cortezas superficiales.	6
	Un relieve desigual, con laderas escarpadas y paisajes muy diversificados.	5
	Grandes pérdidas de la cubierta forestal a causa de repetidos incendios de bosques.	12
	Condiciones de crisis en la agricultura tradicional, con el consiguiente abandono de tierras y deterioro del suelo y de las estructuras de conservación del agua.	7
	Ocasional explotación insostenible de los recursos hídricos, que es causa de graves daños ambientales, incluidos la contaminación química, la salinización y el agotamiento de los acuíferos.	10
	Concentración de la actividad económica en las zonas costeras como resultado del crecimiento urbano, las actividades industriales, el turismo y la agricultura de regadío.	2

Según los encuestados/as, la principal causa del avance de la desertificación (MODA) es:

*Grandes pérdidas de la cubierta forestal a causa de repetidos incendios de bosques.*

En el apartado de recursos del tema puedes ver este ejemplo en una presentación llamada “**Así se calcula de la moda**”.

Comprueba que lo has entendido

8. En un determinado paraje se ha medido la altura de 11 olivos. Las alturas medidas han sido: 3,5 m; 3 m; 2,8 m; 3,4 m; 3,1 m; 2,6 m; 3,8 m; 3 m; 2,7 m; 2,8 m; 3,3 m. ¿Cuál será la altura media de los once olivos del paraje?

- 3 m.
- 3,4 m.
- 3,1 m.



9. Se realiza una encuesta a 100 personas preguntando si separan o no los residuos para reciclarlos, siendo los resultados los recogidos en la tabla siguiente. ¿Cuál es la moda?



	Nº de respuestas
Siempre, clasificando en las categorías: orgánica, vidrio, envases y papel.	10
Siempre, pero sólo papel y vidrio.	15
Casi siempre el papel	23
Casi siempre el vidrio	18
Normalmente no	16
Nunca	10
Otras opciones	8

- Casi siempre el papel.
- Siempre, clasificando en las categorías: orgánica, vidrio, envases y papel.
- Casi siempre el vidrio.

## Parámetros de dispersión



Los parámetros de dispersión **nos informan sobre lo bien (o lo mal) que la media aritmética representa al conjunto de datos**

Vienen a ser algo así como **indicadores de cómo de agrupados están los datos en torno a la media**. Si están muy agrupados, los parámetros de dispersión tomarán valores pequeños. Pero si no lo están, si están muy "dispersos", tomarán valores más grandes.

**IMPORTANTE: solo pueden calcularse si la variable es cuantitativa**  
 (claro, porque solo entonces hay "media aritmética")

El más sencillo de todos los parámetros de dispersión es el **rango** (también es el menos útil, la verdad).



**El rango es la diferencia entre el dato de mayor valor y el dato de menor valor.**

Por tanto, para calcularlo, solo hay que localizar el dato más grande y el más pequeño y... simplemente restarlos. Puedes ver aquí algunos...

### Ejemplos...



¿Recuerdas el ejemplo de los maceteros? Pues bien, el rango de esa variable estadística es...

$$7 - 2 = 5$$

Porque el dato más grande es 7 y el más pequeño es 2.

¿Y en el caso de las calificaciones de las tareas? La calificación más alta era 9 ¿recuerdas? y la más baja era 4. Por tanto, el rango será en este caso...

$$9 - 4 = 5$$

¡Qué coincidencia!

Uno de los parámetros de dispersión más empleados es la **varianza**. Te la puedes encontrar representada con el símbolo  $S^2$  o bien como  $\sigma^2$ . Para calcularla (a mano) hay que hacer algunas cuentecillas. Observa en el siguiente ejemplo cómo se calcula la varianza.



Por ejemplo...



## Calculamos la varianza

Se ha realizado una encuesta a 50 personas preguntando por el número de parajes o parques naturales que conoce de Andalucía, siendo sus respuestas:

Dato ( $x_i$ )	Frec. Abs. ( $f_i$ )
0	7
1	10
2	16
3	11
4	4
5	2
<b>Total</b>	<b>50</b>



Vamos a calcular la **varianza**. Para ello tenemos dos opciones:

OPCIÓN A: Aplicar la fórmula 
$$S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot f_i}{N}$$

OPCIÓN B: O la fórmula equivalente 
$$S^2 = \frac{\sum_{i=1}^n x_i^2 \cdot f_i}{N} - \bar{x}^2$$

Elegimos la opción B

1ª. Elevamos los datos al cuadrado y los sumamos directamente uno a uno

$$\begin{array}{cccccccccccccccccccc}
 0 & + & 1 & + & 4 & + & 1 & + & 0 & + & 9 & + & 16 & + & 16 & + & 9 & + \\
 1 & + & 1 & + & 4 & + & 4 & + & 4 & + & 25 & + & 9 & + & 4 & + & 1 & + \\
 9 & + & 9 & + & 9 & + & 4 & + & 4 & + & 9 & + & 16 & + & 0 & + & 4 & + \\
 9 & + & 25 & + & 4 & + & 1 & + & 0 & + & 1 & + & 4 & + & 1 & + & 0 & + \\
 16 & + & 4 & + & 4 & + & 9 & + & 0 & + & 0 & + & 1 & + & 4 & + & 9 & + \\
 4 & + & 4 & + & 9 & + & 4 & + & 1 & = & 287
 \end{array}$$

Se divide entre el número de entrevistas: 
$$\frac{287}{50} = 5.74$$

Al resultado anterior se le resta la media aritmética elevada al cuadrado

$$5.74 - (2.02)^2 = 5.74 - 4.0804 = 1.6596$$

Por tanto la **varianza** es **1.6596**

2ª. También podemos hacerlo con la ayuda de la tabla de frecuencias

Dato ( $x_i$ )	Frec. Abs. ( $f_i$ )	$x_i^2 \cdot f_i$
0	7	0
1	10	10
2	16	64
3	11	99
4	4	64
5	2	50
	<b>50</b>	<b>287</b>

Para rellenar esta nueva columna, multiplicamos cada dato por si mismo y por su frecuencia absoluta

Ejemplo:  $2 \cdot 2 \cdot 16 = 64$

Sumamos toda la columna y el resultado es la suma total de datos

Se divide entre el número de entrevistas:  $\frac{287}{50} = 5.74$

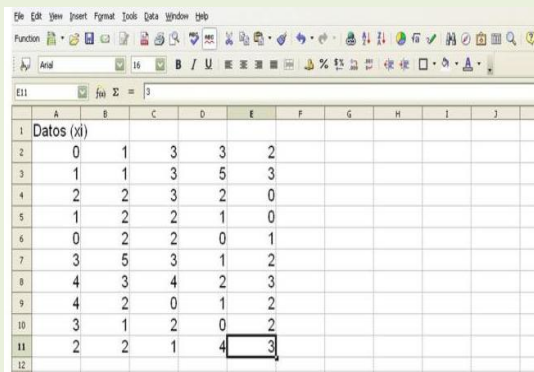
Al resultado anterior se le resta la media aritmética elevada al cuadrado

$$5.74 - (2.02)^2 = 5.74 - 4.0804 = 1.6596$$

Por tanto la varianza es **1.6596**

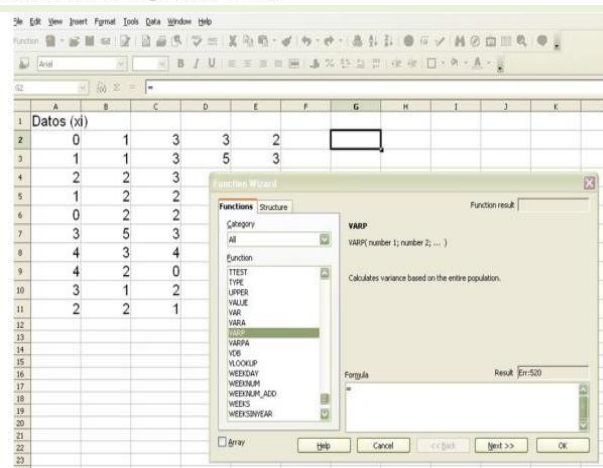
3ª. Usamos una hoja de cálculo (tipo Open Office Cal o Excel)

**Paso 1** Introducimos los datos uno a uno en una hoja de cálculo, cada dato ocupará una casilla



	A	B	C	D
1	Datos (xi)			
2	0	1	3	2
3	1	1	3	5
4	2	2	3	0
5	1	2	2	1
6	0	2	2	0
7	3	5	3	1
8	4	3	4	2
9	4	2	0	1
10	3	1	2	0
11	2	2	1	4

**Paso 2** Elegimos una casilla en blanco y elegimos la acción INSERT y después FUNCTION. De la lista de funciones, elegimos **VARP**



### Paso 3

Pulsamos el botón NEXT y seleccionamos con el ratón todas las casillas en las que estén escritos los datos, después aceptamos

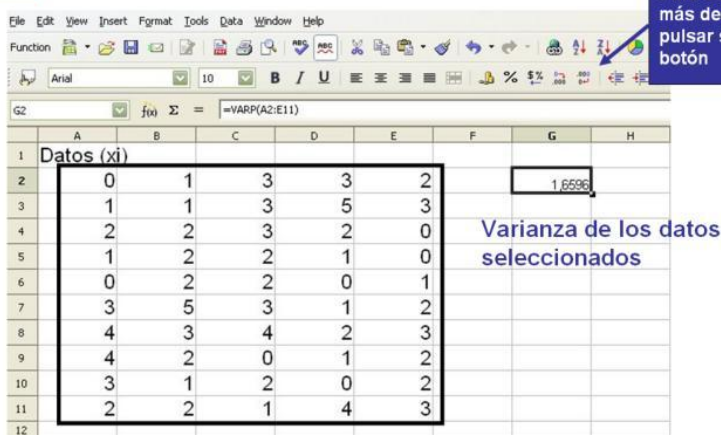
También podemos escribir directamente **=VARP(A2:E11)**



### Paso 4

En la casilla seleccionada aparecerá el valor de la varianza

Para obtener más decimales pulsar sobre el botón



Varianza de los datos seleccionados

La fórmula general para calcular la varianza es:

$$S^2 = \frac{\sum_{i=1}^n x_i^2 \cdot f_i}{N} - \bar{x}^2$$

Diagram explaining the formula components:

- Sumatoria**: Points to the summation symbol  $\sum$ .
- Del producto del cuadrado de los datos por su frecuencia**: Points to  $x_i^2 \cdot f_i$ .
- El cuadrado de la media aritmética**: Points to  $\bar{x}^2$ .
- Menos**: Points to the minus sign  $-$ .
- Dividido**: Points to the division line  $/$ .
- Tamaño de la muestra**: Points to  $N$ .
- Varianza**: Points to  $S^2$ .

En el apartado de recursos del tema puedes ver el ejemplo anterior en una presentación llamada *“Así se calcula de la varianza”*.



La varianza, en sí, tampoco es un parámetro demasiado útil; no nos da una información fácil de interpretar. Pero sí es importante calcularla, porque a partir de ella podremos calcular la **desviación típica** y, ésta... sí que es útil.

La desviación típica se suele representar con los símbolos  $S$  o  $\sigma$  y calcularla es muuuuy fácil. Si te fijas bien en los símbolos, te darás cuenta de que...



**La desviación típica es la raíz cuadrada de la varianza**

$$S = \sqrt{S^2}$$

Observa en este ejemplo cómo se calcula...

Por ejemplo...



### Calculamos la desviación típica

Se ha realizado una encuesta a 50 personas preguntando por el número de parajes o parques naturales que conoce de Andalucía, siendo sus respuestas:

Vamos a calcular la **desviación típica**.

Para ello tenemos que calcular antes la varianza y luego hacer la raíz cuadrada.

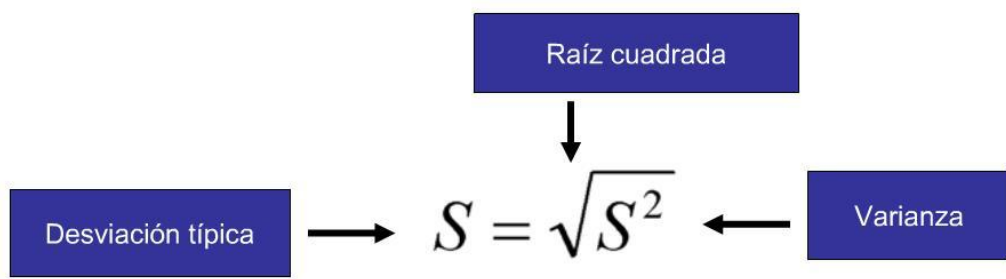
$$S = \sqrt{S^2}$$

En el ejemplo anterior ya calculamos la varianza; ahora tan solo tenemos que hacer la raíz cuadrada al número que obtuvimos entonces:

$$\sqrt{1.6596} \approx 1.2883$$

Si hubiésemos empleado una hoja de cálculo (tipo Excel u Open Office Calc), podríamos haber obtenido directamente la desviación típica sin calcular antes la varianza. Solo tendríamos que haber empleado la función **STDEVP**

RECUERDA... para calcular la **desviación típica**...



En el apartado de recursos del tema puedes ver el ejemplo anterior en una presentación llamada "**Así se calcula de la desviación típica**".

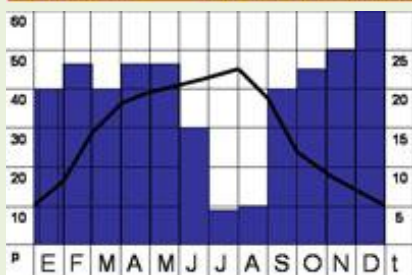
La desviación típica es el **parámetro de dispersión más utilizado**, y sí que tiene un significado muy claro. Nos informa sobre "la distancia media" entre los datos y la media aritmética. **Cuanto menor sea la desviación típica, más representativa será la media**, puesto que los datos estarán más agrupados a su alrededor.

### Por ejemplo...



Imagina que en dos poblaciones andaluzas realizan un estudio sobre las temperaturas máximas que han registrado durante el mes de julio, y que los resultados obtenidos (en °C) en cada población son los de la tabla siguiente:

Villabochoorno	27	28	30	30	30	30	31	32	28	28	28	29	29	30	32	31	32	34	31	29	30	27	27	29	30	33	32	31	30	32	30
Torrecazor	30	27	34	38	40	41	34	32	34	25	22	21	21	25	30	34	33	33	30	27	25	23	23	24	32	32	34	33	33	31	29



La temperatura media en ambos pueblos ha sido la misma, 30°C (compruébalo y así vas practicando un poco). Sin embargo, ese dato no es igual de significativo en ambos casos; en uno de los pueblos la temperatura es "más constante", están menos "dispersas"... ¿en cuál?

Con tan poquitos datos, la respuesta se puede dar con un vistazo atento a los mismos, pero si tuviésemos miles de datos... la cosa sería distinta. Tendríamos, en ese caso, que acudir a la desviación típica. En un pueblo la desviación típica no llega a 2°C,



¿Sabrías decir a qué pueblo corresponde cada dato? Merece la pena que calcules las desviaciones típicas de las dos series de datos... te servirá como práctica.

Por último, otro parámetro muy utilizado es el **coeficiente de variación** (C.V). Para calcularlo necesitamos saber antes tanto la media aritmética como la desviación típica.



El coeficiente de variación se calcula con la fórmula...

$$C.V. = \frac{\text{desviación típica}}{\text{media aritmética}} \times 100$$

El coeficiente de variación es un número relativo y **se expresa en %**. Es muy útil porque, como la desviación típica, nos informa sobre "lo agrupados" que están los datos alrededor de la media. Pero además, **nos permite comparar** estudios realizados, por ejemplo, en diferentes unidades o cuyas medias sean diferentes.

### Por ejemplo...



Si en un estudio sobre la temperatura en cierto mes en dos ciudades, los resultados hubiesen sido:

CIUDAD	Temperatura media (°C)	Desviación típica (°C)
Joquefresco	7	2
Agustitown	25	5

Podríamos estar tentados a pensar que el estudio hecho en Agustitown es peor que el de Joquefresco, ya que su desviación típica es más grande. Sin embargo, si lo piensas bien, observa que no es lo mismo "equivocarse" en 5 cm cuando mides la longitud de algo que mide 25 cm, que "equivocarse" en 2 cm cuando mides algo que mide solo 7 cm. Pues algo parecido sucede aquí.

Si calculamos el coeficiente de variación de cada estudio...

$$C.V. = \frac{2}{7} \times 100 = 28,5\% \quad C.V. = \frac{5}{25} \times 100 = 20\%$$

Podemos ver que, en realidad, el estudio de Agustitown es "mejor" que el de Joquefresco. En términos relativos, las temperaturas en Agustitown se agrupan más en torno a la media que en Joquefresco.



## Comprueba que lo has entendido

10. En un estudio sobre los metros cuadrados ocupados por las distintas zonas verdes en dos localidades, se recogieron los datos que ves en la tabla de la derecha.

¿Cuál de las dos localidades presenta una distribución de zonas verdes más "dispersa"? (Para responder, tendrás que calcular el coeficiente de variación de los metros cuadrados destinados a zona verde de ambas localidades)

- La localidad 1.
- La localidad 2.
- Ambas por igual.



Localidad 1		Localidad 2	
Zona	m <sup>2</sup>	Zona	m <sup>2</sup>
A	780	A	4500
B	1080	B	600
C	2200	C	1800
D	2800	D	5400
E	5600	E	1000
F	950	F	700
G	4200	G	1900
H	2600	H	6100
I	4100		
J	3500		



## Curiosear...

Las calculadoras científicas permiten hacer cálculos estadísticos. Investiga en tu calculadora científica (si es que tienes una) o en la de tu ordenador (o en el de tu clase), e intenta localizar los símbolos de los parámetros estadísticos.

## Para saber más...



Si quieres tener una información completa sobre las **estadísticas oficiales sobre la situación medioambiental** tanto a nivel andaluz como de todo el territorio español, puedes visitar los enlaces que encontrarás en el apartado de recursos web del tema:

- **Estadísticas de la Consejería de Medio Ambiente.**
- **Estadísticas del Ministerio de Medio Ambiente.**
- **Instituto Nacional de Estadística.**



## Comprueba que lo has entendido (soluciones)

1. Las tres respuestas son correctas.

- a. Dos personas en un club náutico es muy probable que sean aficionados al buceo. Es evidente que no pueden representar a todos los españoles.
- b. La compañía eléctrica es, probablemente, la dueña de las centrales térmicas. Por esa razón a ella le conviene que el resultado del estudio sea el que publicita.
- c. Para eso no hace falta un estudio estadístico. Todo el mundo sabe que todo el mundo respira ¿no?

2. La respuesta correcta es la **c**, y no necesita explicación. Es evidente que se trata de una magnitud física y los valores que toma deben ser medidos con un instrumento de medida adecuado.

3. Todo depende de si la variable se puede expresar con un número o no.

Energía aportada por distintas marcas de muesli.	CUANTITATIVA
Sistema de calefacción utilizado en el invierno por familias de Madrid.	CUALITATIVA
Volumen de basura generado por las familias de una barriada de Granada.	CUANTITATIVA
Soluciones al problema de la contaminación de las aguas.	CUALITATIVA

4. La respuesta correcta es la **c**.

Es muy probable que las 50 personas que están en el polideportivo estén encantadas con él. Su opinión no será representativa de la opinión del barrio en general.

Es muy probable que mis amistades tengan gustos similares a los míos, y puede que muchas de ellas ni siquiera sean del barrio. No son representativas para el estudio que queremos hacer.

Haciendo la compra por la mañana en el mercado encontraremos sobre todo mujeres de mediana o avanzada edad; un grupo que no representa a toda la población del barrio.

5. La respuesta correcta es la **b**.

Una sola muestra no será significativa, porque el resultado dependerá de cuándo y dónde se ha tomado.

Tanto si tomamos las muestras donde la fábrica vierte sus desechos como en el nacimiento del río, obtendremos (con total probabilidad) aguas muy contaminadas y aguas muy puras, respectivamente. En ninguno de los dos casos la muestra será representativa del agua de todo el cauce.

6. La respuesta correcta es la **opción b**, porque si contamos el número de animales de cada tipo tenemos:

paloma, gorrión, gato, perro, koi, ardilla, hormiga, mosquito, mosca, araña, cigüeña, goldfish, mirlo, avispa, rana, lagartija, salamanquesa, gusano, ratón, topo, urraca, golondrina, libélula, carpa, niños, grillo, escarabajo, cochinilla, pato, cisne

Aves	paloma, gorrión, cigüeña, mirlo, urraca, golondrina, pato, cisne.	8
Peces	koi, goldfish, carpa.	3
Mamíferos	gato, perro, ardilla, ratón, topo, niños.	6
Reptiles	lagartija, salamanquesa.	2
Anfibios	rana	1
Invertebrados	hormiga, mosquito, mosca, araña, avispa, gusano, libélula, grillo, escarabajo, cochinilla.	10

7. La respuesta correcta es la **c**.

La primera gráfica no puede ser porque, por ejemplo, el sector que representa a los medicamentos usados es más grande que el que representa a la ropa y residuos textiles. La tabla de datos, sin embargo, muestra que las toneladas de medicamentos usados (610) son menos que las de ropa y residuos textiles (810)

Tampoco la segunda gráfica, el diagrama de barras, puede ser. La altura de las barras no tiene nada que ver con los datos de la tabla. Por ejemplo, a los aceites usados, que en la tabla corresponden más de 14500 toneladas, se le asigna una barra cuya altura marca menos de 10000 toneladas.

8. Para responder debemos sumar todos los datos (son pocos, solo once) y luego dividir entre el número de datos:

$$media = \frac{3,5 + 3 + 2,8 + 3,4 + 3,1 + 2,6 + 3,8 + 3 + 2,7 + 2,8 + 3,3}{11} = 3,0909 \cong 3,1$$

Por tanto, redondeando, la altura media de los olivos resulta ser de 3,1 m. La respuesta correcta será la **c**.

9. La moda es el dato o la respuesta más frecuente, la que más se repite. En este caso la respuesta que más se repite es la de "Casi siempre el papel", con 23 repeticiones. Por tanto, esa es la moda y la respuesta correcta sería la **a**.
10. Tenemos que calcular el coeficiente de variación para cada una de las localidades. Para ello es necesario conocer previamente la media y la desviación típica y, para calcular esta última necesitamos también la varianza. Pero también podemos usar una hoja de cálculo para hacerlo. Esta última opción es la que elegimos.

Localidad 1	
Zona	m <sup>2</sup>
A	780
B	1080
C	2200
D	2800
E	5600
F	950
G	4200
H	2600
I	4100
J	3500

Localidad 2	
Zona	m <sup>2</sup>
A	4500
B	600
C	1800
D	5400
E	1000
F	700
G	1900
H	6100

LOCALIDAD 1	
Media:	2781
Varianza:	2537521,11
Desviación típica:	1592,95986
Coeficiente de variación:	57,28%

LOCALIDAD 2	
Media:	2750
Varianza:	4974285,71
Desviación típica:	2230,31068
Coeficiente de variación:	81,10%

A la vista de los resultados, se deduce que la localidad 2 tiene sus zonas verdes "más dispersas" en cuanto a su extensión, es decir, las zonas verdes de la localidad 1 son de extensión más parecida entre sí que las de la 2.